On a relationship between Laplacian eigenmaps and diffusion maps.

Minh Tang[1] Michael Trosset[2]

[1]School of Informatics and Computing
Indiana University, Bloomington

[2]Department of Statistics
Indiana University, Bloomington

This work was funded by a grant from the Office of Naval Research.

Problem Description

We consider the problem of constructing a low-dimensional Euclidean representation of data described by pairwise similarities.

The low-dimensional representation can served as the basis for other exploitation tasks, e.g., visualization, clustering, or classification.

Our basic strategy is:

1. Transform the similarities into some notion of dissimilarities;
2. Embed the derived dissimilarities, e.g., by classical multidimensional scaling [Torgerson(1952)], [Gower(1966)].

Our concerns are closely related to the concerns of manifold learning. Some manifold learning techniques can be interpreted as transformations from similarities to dissimilarities.

Some Terminologies

A dissimilarity matrix $\mathbf{\Delta} = (\delta_{ij})$ is a hollow, symmetric, non-negative matrix. Larger values indicate that the objects are more dissimilar.

A similarity matrix $\mathbf{\Gamma} = (\gamma_{ij})$ is a symmetric, non-negative matrix. Larger values indicate that the objects are more similar.

A $n \times n$ dissimilarity matrix $\mathbf{\Delta} = (\delta_{ij})$ is a Type-2 Euclidean distance matrix (EDM-2) iff there some $x_1, \dots, x_n \in \mathbb{R}^p$ such that $\delta_{ij} = \|x_i - x_j\|^2$.

There is an equivalence between EDM-2 and psd matrices.

- If $\mathbf{\Delta}$ is EDM-2, then $\mathbf{B} = \tau(\mathbf{\Delta}) = -\frac{1}{2}\mathbf{P}\mathbf{\Delta}\mathbf{P}$ is psd, where $\mathbf{P} = (\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)$.
- If $\mathbf{B}$ is psd then $\mathbf{\Delta} = \kappa(\mathbf{B}) = \mathrm{diag}(\mathbf{B})\mathbf{1}\mathbf{1}^T - 2\mathbf{B} + \mathbf{1}\mathbf{1}^T\mathrm{diag}(\mathbf{B})$ is EDM-2.

Isomap

Isomap [Tenenbaum et al.(2000)] is one of the best known manifold learning algorithm. Suppose that $y_1, y_2, \ldots, y_n \in \mathbb{R}^q$ lie on a d-dimensional manifold. To represent $y_1, y_2, \ldots, y_n$ as $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$, Isomap replaces Euclidean distance in $\mathbb{R}^q$ with a clever approximation of geodesic distance on the manifold as follows:

1. Replace Euclidean distance with approximate geodesic distance.
   - (a) Construct a weighted graph $G = (V, E, \omega)$ with n vertices. Fix some $\epsilon \geq 0$ and let $v_i \sim v_j$ iff $\|y_i - y_j\| \leq \epsilon$. If $v_i \sim v_j$, set $\omega_{ij} = \|y_i - y_j\|$.
   - (b) Compute $\boldsymbol{\Delta} = (\delta_{ij})$ where δ_{ij} is the shortest path distance between v_i and v_j in G.
2. Embed $\boldsymbol{\Delta}$ by CMDS.

From Similarities to Distances on Graphs

The Isomap recipe can be adapted to work with similarities as follows.

Given a $n \times n$ similarities matrix $\mathbf{\Gamma} = (\gamma_{ij})$:

1. Transform the similarities to distances.
   (a) Construct a weighted graph $G = (V, E, \omega)$ with n vertices and edge weights $\omega_{ij} = \gamma_{ij}$.
   (b) Construct a matrix $\mathbf{\Delta} = (\delta_{ij})$ that measures some suitable distance on G.
2. Embed $\mathbf{\Delta}$.

Several popular approaches to transform from similarities to distances relies on the concept of a random walk.

Assume that G is connected. Let $\mathbf{s} = \mathbf{\Gamma}\mathbf{1}$ and $\mathbf{S} = \mathrm{diag}(\mathbf{s})$. Then the random walk on $G = (V, E, \omega)$ is the Markov chain with state space V and transition probabilities $\mathbf{P} = \mathbf{S}^{-1}\mathbf{\Gamma}$. The stationary distribution $\boldsymbol{\pi}$ of $\mathbf{P}$ exists and is unique, and furthermore, $\lim_{k\to\infty} \mathbf{P}^k = \mathbf{1}\boldsymbol{\pi}^T := \mathbf{Q}$.

Expected Commute Time

Following [Kemeny and Snell(1960)], let

$$\mathbf{\Pi} = \mathrm{diag}(\boldsymbol{\pi}) \quad \text{and} \quad \mathbf{Z} = (\mathbf{I} - \mathbf{P} + \mathbf{Q})^{-1}.$$

The expected first passage times are given by

$$\mathbf{M} = (\mathbf{1}\mathbf{1}^T \mathrm{diag}(\mathbf{Z}) - \mathbf{Z})\mathbf{\Pi}^{-1}$$

and the expected commute times are

$$\mathbf{\Delta}_{\mathrm{ect}} = \mathbf{M} + \mathbf{M}^T = \kappa(\mathbf{Z}\mathbf{\Pi}^{-1})$$

It turns out that $\mathbf{Z}\mathbf{\Pi}^{-1} \succeq 0$. $\mathbf{\Delta}_{\mathrm{ect}}$ is thus EDM-2.

Diffusion Distances

Let $\mathbf{e}_i$ and $\mathbf{e}_j$ denote point masses at vertices v_i and v_j. After r time steps, under the random walk model with transition matrix $\mathbf{P}$, these distributions had diffused to $\mathbf{e}_i^T \mathbf{P}^r$ and $\mathbf{e}_j^T \mathbf{P}^r$.

The diffusion distance [Coifman and Lafon(2006)] at time r between v_i and v_j is

$$\rho_r(v_i, v_j) = \|\mathbf{e}_i^T \mathbf{P}^r - \mathbf{e}_j^T \mathbf{P}^r\|_{1/\boldsymbol{\pi}}$$

where the inner product $\langle \cdot, \cdot \rangle_{1/\boldsymbol{\pi}}$ is defined as

$$\langle \mathbf{u}, \mathbf{v} \rangle_{1/\boldsymbol{\pi}} = \sum_k u(k) v(k) / \pi(k)$$

It turns out that $\Delta_{\rho_r^2} = \kappa(\mathbf{P}^{2r} \mathbf{\Pi}^{-1})$. Because $\mathbf{P}^{2r} \mathbf{\Pi}^{-1} \succeq 0$, $\Delta_{\rho_r^2}$ is EDM-2.

Some Remarks on ECT and Diffusion Distances

1. $\mathbf{\Delta}_{\mathrm{ect}}$ can be written as

$$\mathbf{\Delta}_{\mathrm{ect}} = \kappa(\mathbf{Z}\mathbf{\Pi}^{-1}) = \kappa\Big(\sum_{k=0}^{\infty}(\mathbf{P} - \mathbf{Q})^k \mathbf{\Pi}^{-1}\Big).$$

   The expected commute time between v_i and v_j take into account paths of all length between v_i and v_j.

2. Even though $(\mathbf{P} - \mathbf{Q})^k = \mathbf{P}^k - \mathbf{Q}$ for $k \geq 1$, $\mathbf{Q}\mathbf{\Pi}^{-1} = \mathbf{1}\mathbf{1}^T$ and $\kappa(\mathbf{1}\mathbf{1}^T) = \mathbf{0}$, one cannot write $\mathbf{\Delta}_{\mathrm{ect}} = \kappa\Big(\sum_{k=0}^{\infty} \mathbf{P}^k \mathbf{\Pi}^{-1}\Big)$ because $\sum_{k=0}^{\infty} \mathbf{P}^k \mathbf{\Pi}^{-1}$ doesn't necessarily converge.

3. $\mathbf{\Delta}_{\rho_r^2} = \kappa(\mathbf{P}^{2r}\mathbf{\Pi}^{-1}) = \kappa\big((\mathbf{P} - \mathbf{Q})^{2r}\mathbf{\Pi}^{-1}\big)$. Diffusion distance between v_i and v_j at time r take into account only paths of length $2r$.

General Framework for Euclidean Distances on Graphs

We now introduce a general family of Euclidean distances constructed from random walks on graphs.

Let f be a real-valued function with a series expansion

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots$$

and radius of convergence $R \geq 1$.

If $f(x) \geq 0$ for $x \in (-1, 1)$ (and $\mathbf{P}$ is irreducible and aperiodic), then

$$\boldsymbol{\Delta} = \kappa(f(\mathbf{P} - \mathbf{Q})\boldsymbol{\Pi}^{-1}) = \kappa\Big((a_0\mathbf{I} + a_1(\mathbf{P} - \mathbf{Q}) + a_2(\mathbf{P} - \mathbf{Q})^2 + \cdots)\boldsymbol{\Pi}^{-1}\Big)$$

is well-defined and EDM-2. In the above equation, f acts on the matrix $\mathbf{P} - \mathbf{Q}$ and not on the entries of $\mathbf{P} - \mathbf{Q}$.

Euclidean Distances on Graphs: Some Examples

$$\mathbf{\Delta} = \kappa(f(\mathbf{P}-\mathbf{Q})\mathbf{\Pi}^{-1}) = \kappa\Big((a_0\mathbf{I} + a_1(\mathbf{P}-\mathbf{Q}) + a_2(\mathbf{P}-\mathbf{Q})^2 + \cdots)\mathbf{\Pi}^{-1}\Big)$$

The following functions generate $\mathbf{\Delta}$ that are EDM-2.

- $f(x) = 1/(1-x)$ gives expected commute time.
- $f(x) = 1/(1-x)^2$ gives a distance that, in comparison to expected commute time, assign longer paths higher weights.
- $f(x) = x^{2r}$ gives diffusion distance at time r.
- $f(x) = -\log(1-x^2)$ gives a distance that take into account only paths of even lengths, with longer paths having lower weights.
- $f(x) = \exp(x)$ gives a distance that take into account paths of short length only, i.e. long paths have almost no weights.

Embedding $\mathbf{\Delta} = \kappa(f(\mathbf{P} - \mathbf{Q})\mathbf{\Pi}^{-1})$ in $\mathbb{R}^d$: Method 1

Embed $\mathbf{\Delta}$ by classical MDS.

1. Compute

$$\mathbf{B} = \tau(\mathbf{\Delta}) = -\frac{1}{2}(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)f(\mathbf{P} - \mathbf{Q})\mathbf{\Pi}^{-1}(\mathbf{I} - \mathbf{1}\mathbf{1}^T/n)$$

2. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ denote the eigenvalues of $\mathbf{B}$ and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ denote the corresponding set of orthonormal eigenvectors. Then

$$\mathbf{X} = \left[\sqrt{\lambda_1}\mathbf{v}_1 | \sqrt{\lambda_2}\mathbf{v}_2 | \cdots | \sqrt{\lambda_d}\mathbf{v}_d\right]$$

produces a configuration of points in $\mathbb{R}^d$.

Embedding $\boldsymbol{\Delta} = \kappa(f(\mathbf{P} - \mathbf{Q})\boldsymbol{\Pi}^{-1})$ in $\mathbb{R}^d$: Method 2

Embed $\boldsymbol{\Delta}$ by the eigenvalues and eigenvectors of $\mathbf{P}$.

1. Let $\mu_1, \mu_2, \ldots, \mu_{n-1}$ be the eigenvalues of $\mathbf{P}$, sorted so that $f(\mu_i) \geq f(\mu_{i+1})$ and $\mu_i \neq 1$ for $1 \leq i \leq n-1$, and let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n-1}$ denote the corresponding set of eigenvectors, orthonormal with respect to the inner product $\langle \mathbf{u}, \mathbf{v} \rangle_{\boldsymbol{\pi}} = \sum_k u(k)v(k)\pi(k)$.
2. Then
$$\mathbf{X} = \left[\sqrt{f(\mu_1)}\mathbf{u}_1 | \sqrt{f(\mu_2)}\mathbf{u}_2 | \cdots | \sqrt{f(\mu_d)}\mathbf{u}_d \right]$$
produces a configuration of points in $\mathbb{R}^d$.

Comparing the Embeddings

Method 1: Classical MDS

1. The embedding $\mathbf{X} = \left[\sqrt{\lambda_1}\mathbf{v}_1 | \cdots | \sqrt{\lambda_{n-1}}\mathbf{v}_{n-1}\right]$ recovers $\boldsymbol{\Delta}$ completely.
2. The embedding dimension of $\boldsymbol{\Delta}$ is $n-1$ with probability 1.
3. The best (least squares) d-dim representation of $\mathbf{X}$ is $\mathbf{X}_d = \left[\sqrt{\lambda_1}\mathbf{v}_1 | \cdots | \sqrt{\lambda_d}\mathbf{v}_d\right]$.
4. $\mathbf{X}_d\mathbf{X}_d^T$ is the best rank-d approximation of $\mathbf{B}$.

Method 2: Eigensystem of $\mathbf{P}$

1. The embedding $\mathbf{X} = \left[\sqrt{f(\mu_1)}\mathbf{u}_1 | \cdots | \sqrt{f(\mu_{n-1})}\mathbf{u}_{n-1}\right]$ recovers $\boldsymbol{\Delta}$ completely.
2. The embedding dimension of $\boldsymbol{\Delta}$ is $n-1$ with probability 1.
3. The best (least squares) d-dim representation of $\mathbf{X}$ is (usually) not $\mathbf{X}_d = \left[\sqrt{f(\mu_1)}\mathbf{u}_1 | \cdots | \sqrt{f(\mu_d)}\mathbf{u}_d\right]$
4. Embeddings for different f are (non-uniform) scaling of one another.

Examples of Embeddings

- Diffusion maps [Coifman and Lafon(2006)] is the embedding of diffusion distances using the eigenvalues and eigenvectors of $\mathbf{P}$. The d dimensional embedding is

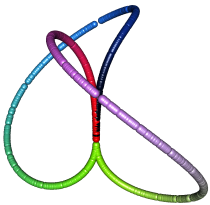
$$\left[\mu_1^r \mathbf{u}_1 | \mu_2^r \mathbf{u}_2 | \cdots | \mu_d^r \mathbf{u}_d\right]$$

- The embedding of expected commute time by CMDS turns out to be equivalent to embedding using the eigenvalues and eigenvectors of the combinatorial Laplacian. If $0 = \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n$ are the eigenvalues of the combinatorial Laplacian $\mathbf{L}$, and $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are the corresponding eigenvectors, then

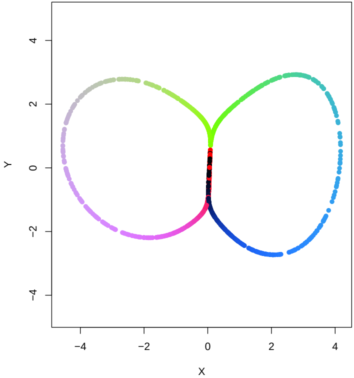
$$C\left[\frac{\mathbf{v}_2}{\lambda_2} \Big| \frac{\mathbf{v}_3}{\lambda_3} \Big| \cdots \Big| \frac{\mathbf{v}_{d+1}}{\lambda_{d+1}}\right]$$

is the d dimensional embedding of $\boldsymbol{\Delta}_{\mathrm{ect}}$, where C is a constant.

Embedding of a 3D curve



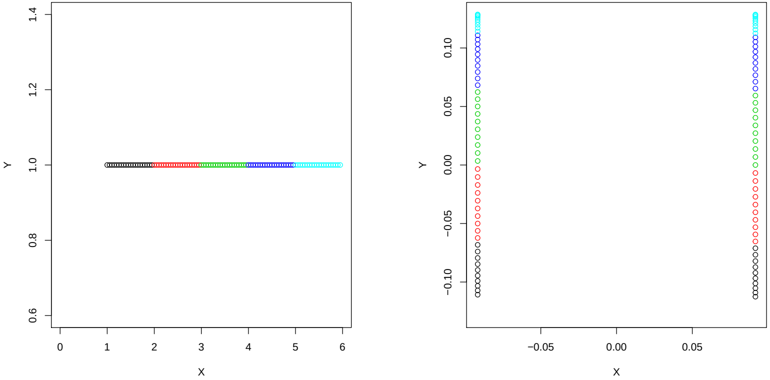
(a)



(b)

Figure: Similarity is computed by $\gamma_{ij} = \exp(-\|x_i - x_j\|^2/\sigma)$, $\sigma = 0.4$.

Paths of Even Length & Diffusion Distances



(a) (b)

R. Coifman and S. Lafon.
Diffusion maps.
Applied and Computational Harmonic Analysis, 21:5–30, 2006.

J. C. Gower.
Some distance properties of latent root and vector methods used in multivariate analysis.
Biometrika, 53:325–338, 1966.

J. G. Kemeny and J. L. Snell.
Finite Markov Chains.
Springer, 1960.

J. B. Tenenbaum, V. de Silva, and J. C. Langford.
A global geometric framework for nonlinear dimensionality reduction.
Science, 290:2319–2323, December 2000.

W. S. Torgerson.
Multidimensional scaling: I. Theory and method.
Psychometrika, 17:401–419, 1952.